Chemometrics and Intelligent Laboratory Systems, 14 (1992) 397–412
Elsevier Science Publishers B.V., Amsterdam

Multivariate spectrometric image analysis

An illustrative study with two constructed examples of metal ions in solution

Fredrik Lindgren * and Paul Geladi

Research Group for Chemometrics, Department of Organic Chemistry, University of Umeå, S-901 87 Umeå (Sweden)

(Received 2 August 1991; accepted 24 October 1991)

Abstract

Lindgren, F. and Geladi, P., 1992. Multivariate spectrometric image analysis. An illustrative study with two constructed examples of metal ions in solution. *Chemometrics and Intelligent Laboratory Systems*, 14: 397–412.

A general introduction to multivariate image analysis (MIA) as a useful tool in chemistry is given, by presenting the analysis of two artificial examples. A comparison of a spectrophotometer and a MIA system is described. Similarities between the two systems are discussed and the advantages of both systems are highlighted. In earlier publications, a classification of pure solid substances with MIA using their reflectances was introduced. This paper further develops the method by considering two examples where aqueous solutions of the metal ions, Co^{2+}, Ni^{2+} and Cu^{2+}, have been investigated. A chemical introduction to the problem and an explanation of the method is given.

INTRODUCTION

Digital image analysis has been around for some time and has made useful contributions to medicine, optical, electron and ion microscopy, satellite imaging, remote sensing, astronomy, geology, agriculture, etc. An image is a recording of radiation intensity on some medium. For computer representation, there has to be some form of digitization. The most popular setup is the use of a regular grid of square elements, called pixels, where each pixel represents an intensity. The grid itself is usually rectangular and most often square. This description nicely matches the storage and display capabilities of computers. The first attempts of digital image analysis were ways of using the properties of rectangular arrays of numbers. Methods developed were filtering, noise reduction, deconvolution, segmentation and morphological measurements [1–3].

Since the numbers in the image arrays represent radiation intensities, it is easy to give them an interpretation in spectroscopic or spectrometric terms. In this case, it is well known that multivariate analysis is more accurate and robust than univariate analysis. Therefore, it can be an advantage to look at images in many variables.

The field of satellite imaging [4–7] was the first to adopt multivariate imaging, in parallel with airborne imaging from airplanes and heli-

copters. Satellites monitoring the earth's surface in four wavelength bands, and later in seven wavelength bands, have been around for over a decade. The number of bands is still increasing and the demands for multivariate treatment of the data are therefore high. The first attempts to use many bands (wavelengths, variables) in other fields were mainly limited to RGB (red, green and blue) color imaging.

Nowadays, almost all methods of imaging are capable of generating multivariate image data. The number of applications and the types of radiation (or other physical variables) used are huge [8–11]. Hence, there is a need for a very general family of methods for handling multivariate data in order to extract problem defined information from multivariate images, where both the spatial resolution and the spectral resolution are increasing.

The methods and applications presented in this article represent a contribution towards this goal. The principal component analysis of multivariate images that is presented and the examples emphasizing the necessity of meaningful spectrometric content illustratively show the links between multivariate image analysis (MIA) and spectrometric data analysis.

Principal component analysis is a suitable method for studying the geometry of the multivariate space. The method is free of assumptions, and the parameters are extracted without using any spatial information in the image. The possibility of plotting the results in two- and three-dimensional scatter plots is a useful tool for exploratory data analysis and for classifications.

In a previous investigation [12], an experimental setup was designed for spectrometric analysis of solid substances. The samples studied were four green organic substances (in the form of powders), namely rhodamine B, malachite green, DDCI-4 and IR-26. The substances were placed as four piles on black velvet (background) and put in front of the camera. A principal component analysis was calculated on the generated multivariate image and resulted in a four-component model. The substances were discriminated from each other by using feature space segmentation in the score images. Reflection from a solid surface is a complicated nonlinear phenomenon. The experience gained in earlier investigations will hopefully be useful, when introducing the method for analysis of liquid phases.

In the present study, hydrous solutions of metal ions were investigated by multivariate image analysis. The examples illustrate a typical spectrometric analytical application to liquid samples. The first example is a designed series of aqueous mixtures of cobalt (Co) and nickel (Ni) with gradually changing concentrations. In the second example, the experiment is expanded to include also copper (Cu). For each example, eight congruent images were generated at different wavelengths (460–900 nm) to constitute the raw data in a multivariate image analysis. Identification and classifications of the different solutions were done using score plots and feature space segmentation.

MULTIVARIATE IMAGE ANALYSIS IN CHEMISTRY

It is possible to perturb molecular systems using different kinds of electromagnetic radiation. The degree of perturbation can be measured in many ways. The spectrum of electromagnetic radiation is a medium for communication between the molecular systems and the macroscopic world.

In chemistry, these phenomena are widely used in many applications and various forms. Especially, many analytical applications are developed using electromagnetic radiation for determination of structures and concentrations of molecules in a wide variety of samples. The most common techniques are based on measurements of radiation intensity differences before and after contact with a sample. The computed difference is a direct measurement of absorbed energy and thus of the concentration of the absorbing species. This is the basic principle for the spectrophotometer, today routine equipment in all analytical laboratories.

The detectors in most spectrophotometers are photomultipliers. The photomultiplier measures transmitted or reflected illumination of an irradiated sample and records one single average value

of absorbed radiation. The photomultiplier is a sensitive detector with a large spectral resolution, and ideal for measuring homogeneous samples. For heterogeneous samples, where a distribution of concentrations is present, photomultiplier measurements are rather slow. Substituting the photomultiplier with a TV camera may give new and useful advantages. The most interesting one is the quick measurement of spatially resolved intensity.

A sample viewed by a TV camera is read into lines of pixels, for example, 512 lines with 512 pixels in each line. Instead of summarizing the absorption of radiation by one average value, it can be represented by 262 144 (512 × 512) single observations (pixels), a high-resolution image of the sample. The intensity difference in the generated image is due to a varying degree of absorption within the viewed area. Thus, for homogeneous samples the photomultiplier and the TV camera give the same result. For non-homogeneous samples, the TV camera records the structure of the sample so that for example patterns, gradients and clusters can be detected. The size of the viewed area (sample) can vary from microscopic (the resolution of electron microscopes) to macroscopic (satellite images, astronomy). If the wavelengths of the illuminating light are varied when recording many images, a collection of congruent univariate images, at different wavelengths, is created. These images put together into a stack give a multivariate image containing spectral information, a spectrum for each pixel (see Fig. 1).

The possibility of creating multivariate images in chemistry is not limited to electromagnetic radiation spectrometry [13]. Methods like magnetic resonance imaging (MRI), secondary ion mass spectrometry (SIMS) and electron microscopy, give multivariate images of elementary analytical content of a surface, by measuring response to varying pulse sequences, m/e or electron energy loss.

Methods for volume analysis mostly used in the medical field are X-ray tomography (CAT-scan), MRI and positron emission tomography (PET). These methods can also be given a multivariate content leading to a chemical interpretation regarding the object (patient) studied.

MULTIVARIATE IMAGES AND STATISTICAL ANALYTICAL METHODS

Traditional image analysis is concerned with two-dimensional (object × object) types of data structures. This is the classical univariate analysis of grey images. Algorithms for contrast enhancement, noise reduction, texture finding, segmentation, particle morphology, etc. are the most frequently used.

More interesting, and often even unavoidable, is the analysis of multivariate images (object × object × variable). A multivariate image can be

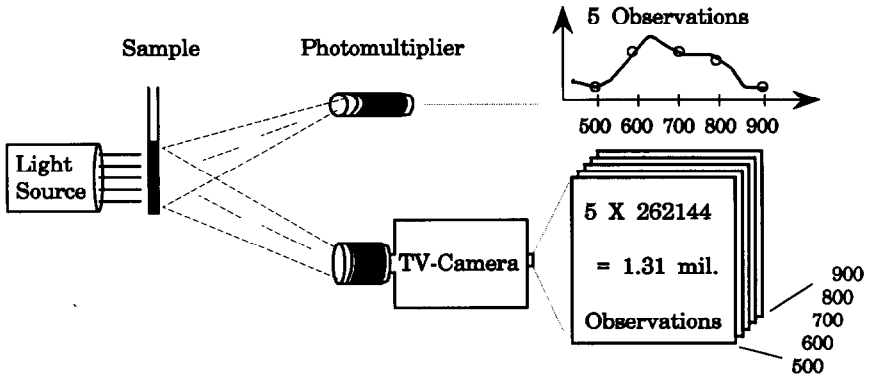

Fig. 1. Two systems of gathering spectral information from an illuminated sample. The photomultiplier has the advantage of having high spectral resolution and sensitivity, but the spatial resolution is low. The TV camera has a high spatial resolution, but problems with sensitivity and with spectral resolution occur. The study presented in this paper takes up some of these problems.

viewed as a stack of univariate images, each image representing one variable. A more useful representation for data analytical purposes is that a multivariate image is an array of pixels (e.g. with two geometrical dimensions), each one associated with a Q-dimensional vector of variables (Q univariate images) [14–16]. In this study the variables are images generated at different wavelengths.

In the field of chemometrics, powerful multivariate statistical tools for chemical applications have been developed during the last decades. Methods such as principal component analysis (PCA) [17], principal component regression (PCR) [18] and partial least squares (PLS) [17–19] are frequently used. These methods have almost exclusively been developed and applied to two-dimensional non-image data structures (objects × variables) [20–22]. Introducing these methods into image analysis gives new possibilities of powerful multivariate image analysis.

To find the systematic structure in the data set, PCA was used. PCA extracts the systematic information in the multi-dimensional space and decomposes it into a few descriptive principal components, orthogonal to each other. Theoretically, PCA corresponds to a mathematical decomposition of the raw data matrix **X** into scores (**T**), also called score images, loadings (**P**) and residuals (**E**) [23–25]. For a geometrical representation, see Fig. 2.

PCA is a data compression method, since the important systematic information is condensed in a few components and the noise part of the data is discarded. The evaluation of the results of a PCA is often done visually, since the sheer amount of numerical results makes any interpretation virtually impossible. Evaluating the score plots, one can find classes of similar objects (pixels) as well as outliers. The loading plots give information of similarities between variables (layers) and show which variables are strongly correlated with the estimated components. Viewing the residuals shows well and badly modelled areas within the image. This localization is typical for multivariate image analysis and does not play a large role in the PCA of most two-dimensional arrays. The context available after analysis in images results and its subjective problem related interpretation are assets for multivariate image analysis.

EXPERIMENTAL

Samples

The chemical substances used in these artificial examples are the metal ions of cobalt (Co^{2+}), nickel (Ni^{2+}) and copper (Cu^{2+}). Products used were cobalt(II) chloride hexahydrate ($CoCl_2 \cdot 6H_2O$, Fluka), cupric chloride dihydrate ($CuCl_2 \cdot 2H_2O$, Analar) and nickel chloride ($NiCl_2$, Vitrum AB). All three chemicals were of the purity denoted purum. Stock solutions of the three metal ions were made and they were calibrated to give equal maximal peak absorption (0.5 Abs) in the spectral range 460–900 nm (Fig. 3). The calibration was carried out with a Shimadzu UV-3101 PC spectrophotometer.

Mixtures of these stock solutions were made according to the scheme outlined in Table 1. Approximately 1 ml of each mixture was later transferred into white plastic cups (20 mm diameter) and placed in front of the camera (Fig. 4).

The circular arrangement of the samples was applied to compensate for camera bias and un-

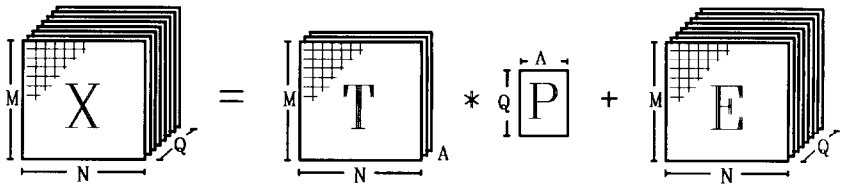

Fig. 2. A data array **X** of size $M \times N \times Q$ is decomposed in A principal components and a residual **E**. If **X** is a multivariate image, Q is the number of wavelength bands and $M \times N$ the image size. The principal component part consists of A score images and A loading vectors.

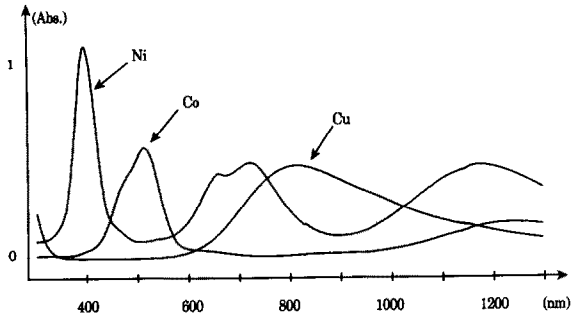

Fig. 3. The absorbance spectra of the three experimental substances, as recorded in a spectrometer with 1 nm wavelength resolution. The spectra overlap substantially in the range 460–900 nm used in the image experiments.

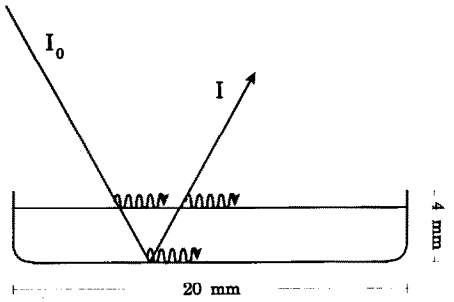

Fig. 4. The arrangement of the solution container. The illumination path is shown. The radiation is fractionated into subvectors both in the surface and on the bottom of the plastic cup.

even illumination in the corners and center (Fig. 5). The background is black velvet, a material with low reflection.

Instrumentation and measurement

The camera used was a DAGE-MTI 70 with a P8029 25.4 mm diameter leddicon (PbS) tube and an AF Nikkor 28 mm 1:2.8 lens. The light source was a Zeiss Superlux with a 300 W xenon lamp connected to a Schott 4 mm fibre optic cable leading to a ring of fibre ends for creating a homogeneous illumination. The optic fibre system was modified for insertion of a filter holder. The filters in the filter holder were Spectron Optik interference filters. The wavelengths used

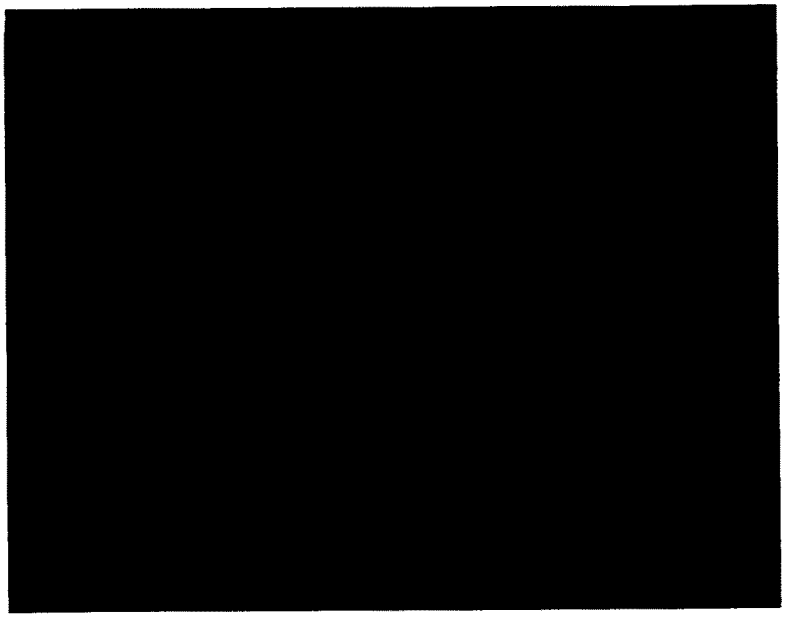

Fig. 5. Example 1, a circular arrangement of the nine plastic cups containing metal ion solutions. This is a color RGB composite image of three randomly chosen univariate images, Nos. 2, 4 and 7. Sample No. 1 is on the top and the other eight samples are arranged in numerical order clockwise. In the picture one can see the sharp reflectance of the illumination fibre ring as ellipsoid-shaped light circles. The yellow squares (50 × 50) show the selected parts of the image that were subjected to further analysis. The image (50 × 50 × 8) from sample No. 1 is placed in the upper left corner and the other eight images are added row-wise in numerical order. See Figs. 6 and 7.

TABLE 1

The percentage content of different metal ions in the samples

Example 1			Example 2			
Sample No.	Co (%)	Ni (%)	Sample No.	Co (%)	Ni (%)	Cu (%)
1	100	0	1	100	0	0
2	95	5	2	0	100	0
3	80	20	3	0	0	100
4	65	35	4	50	50	0
5	50	50	5	50	0	50
6	35	65	6	0	50	50
7	20	80	7	33	33	33
8	5	95				
9	0	100				

were 459, 498, 538, 583, 632, 683, 804 and 897 nm. The band width of the filters is approximately ± 20 nm. Using the filters in the illumination path gives two practical advantages: avoiding the need for extensive refocusing after each time a filter is changed and allowing the use of filters of lower diameter than the camera lens. Digitalization of the video signal was done on a 386/387 PC host with Kontron's IBAS hardware controlled by IBAS 2.0 software. The multivariate image analysis software was written in Flex, a Fortran dialect, using subroutines of the ERDAS toolkit. It operates on a 386/387 PC computer using a Revolution Number Nine image display board and a Multisync Image monitor. Pictures were photographed from this monitor on a 200 ISO color negative film.

All images were digitized in size 512×512 and with integer values in the range 0–255.

TWO EXPLORATORY EXAMPLES

First example

In this work it was tried to broaden the scope of the method to include liquid phases. Illuminating solutions often give a typical light scattering in the surface, causing sharp reflectance. This reflectance often affects the contrast in the image so that useful information is lost. As mentioned earlier, white plastic cups were used as carriers of the solutions. A basic assumption is that the radiation passes through the solution, is reflected at the bottom of the white cup and goes back (Fig. 4). The amount of reflected radiation depends on how great the absorption is. The pathway of the illumination is not easy to explain but a fairly rough estimate is made. The radiation is fractionated into sub-vectors when reaching the sample. Scattering appears probably both in the liquid surface and on the bottom of the cup. Thus, the reflected radiation reaching the camera is the sum of all appearing radiation phenomena. This experimental setup is investigated for two examples.

In the first example, the two compounds, cobalt and nickel, were mixed in a gradient over nine samples (Table 1). The nine cups of sample solutions were arranged in a circular formation in front of the camera. Eight congruent images of size 512×512 were generated in the spectral range 460–900 nm. Every image is an average of ten images. This is done for equalization of the noise of the system. The eight univariate images were collected together into a stack, a multivariate image.

The so-called raw image (Fig. 5) contains large areas of no interest, the background and some sharp circular reflectance from the illumination source. Therefore, areas of interest, yellow in Fig. 5, were selected and cut out as small quadratic pieces ($50 \times 50 \times 8$) from the raw image. These nine sub-images, one for each sample, were later put together to give a new image ($150 \times 150 \times 8$).

It is also possible to see (Fig. 5) an intensity gradient from the middle of the image towards the edges. This phenomenon is mainly caused by illumination errors and biased camera tube.

The eight univariate images constituting the multivariate image are shown in Figs. 6 and 7, together with their wavelengths. Before running a PCA of the three-dimensional matrix, each image in the ($150 \times 150 \times 8$) stack was stretched linearly to fill the whole range of intensities of 0–255. This method is not the same as the usual scaling to variance = 1, but has given satisfactory results in the past. It should be noted here that no

Fig. 6. Example 1, the first four images in the combined ($150 \times 150 \times 8$) multivariate raw image. The wavelengths at which the images were generated, 459, 498, 538 and 583 nm, are given in the figure.

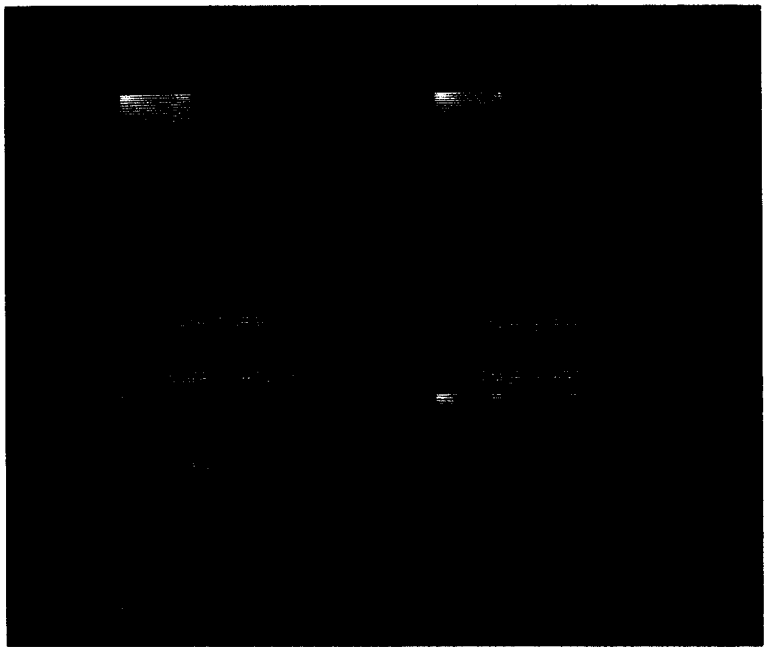

Fig. 7. Example 1, raw image number five to eight generated at 632, 683, 804 and 897 nm.

spectrometric transformation to absorbance or transmission of the image data was used.

A PCA was carried out on the (150 × 150 × 8) pretreated image with mean centering of the variables (images). The variables are in numerical order according to increasing illumination wavelength. The rank of the matrix found was estimated to three. The fourth component contained no information, just noise. Cross-validation [26] was not used, but by a subjective inspection of the PC images a fairly reasonable judgement was made.

The three-component model explained 97.9% of the total sum of squares. Table 2 shows the sum of squares explained by the different components. Fig. 8 shows the three principal component images together with their resulting combination in a RGB composite image. The first principal component is in the blue (B) channel, the second in the green (G) channel and the third one is in the red (R) channel. It is possible to see that the information content decreases and the images get noisier at higher components. The orthogonality aspect of principal components is very clearly seen when comparing the first two score images. The color RGB image shows subjectively that there is some kind of gradient appearing in the image. The images can be viewed but only subjective judgement of their content can be drawn.

TABLE 2

Explained sum of squares for estimated principal components

Example 1		Example 2	
PC No.	SS (%)	PC No.	SS (%)
1	58.7	1	54.0
2	34.1	2	37.1
3	5.1	3	6.4
		4	1.7

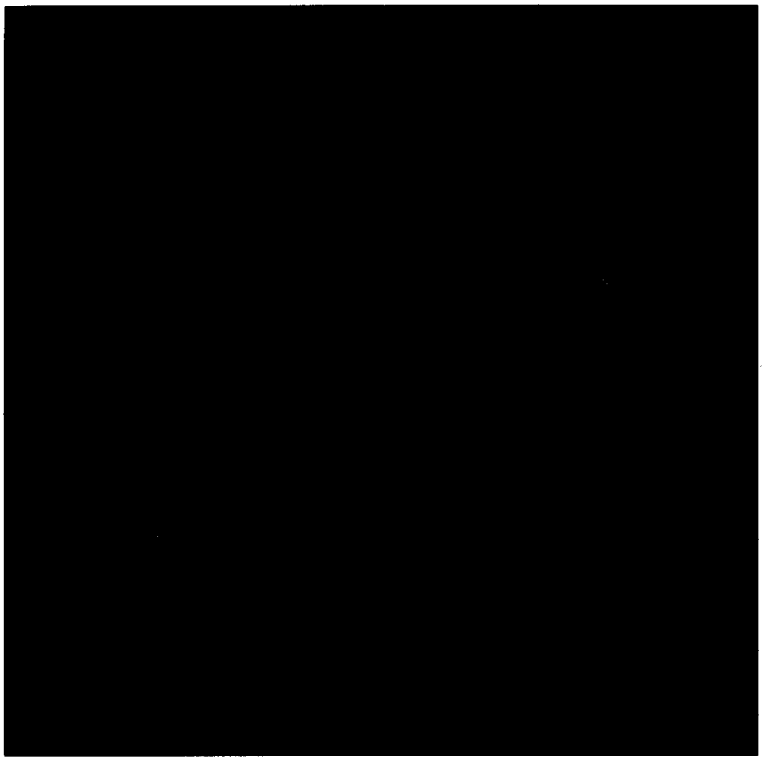

Fig. 8. Example 1, three calculated principal component images and their RGB composite image. The first principal component in the blue channel, the second in the green channel and the third in the red channel.

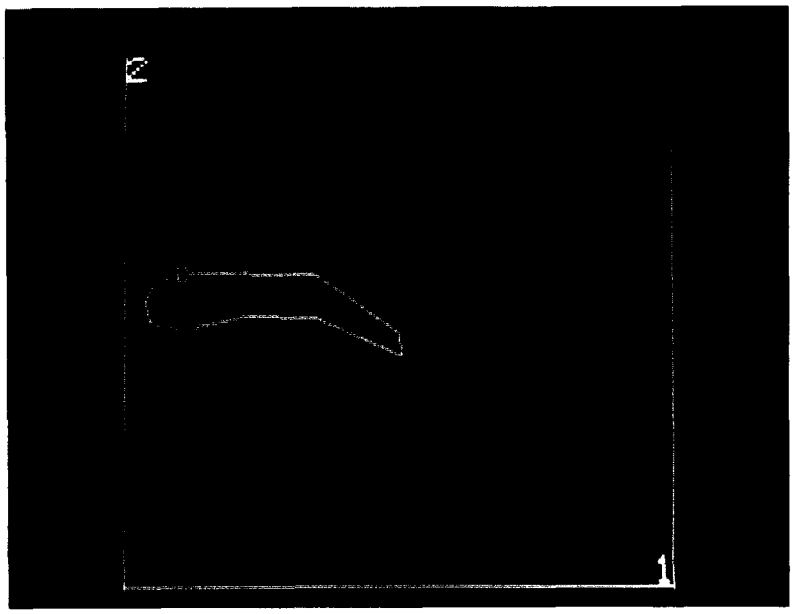

Fig. 9. Example 1, a scatter plot of the second component versus the first. Eight clusters (olive green) appear in the plot. One of the clusters is selected by a score mask.

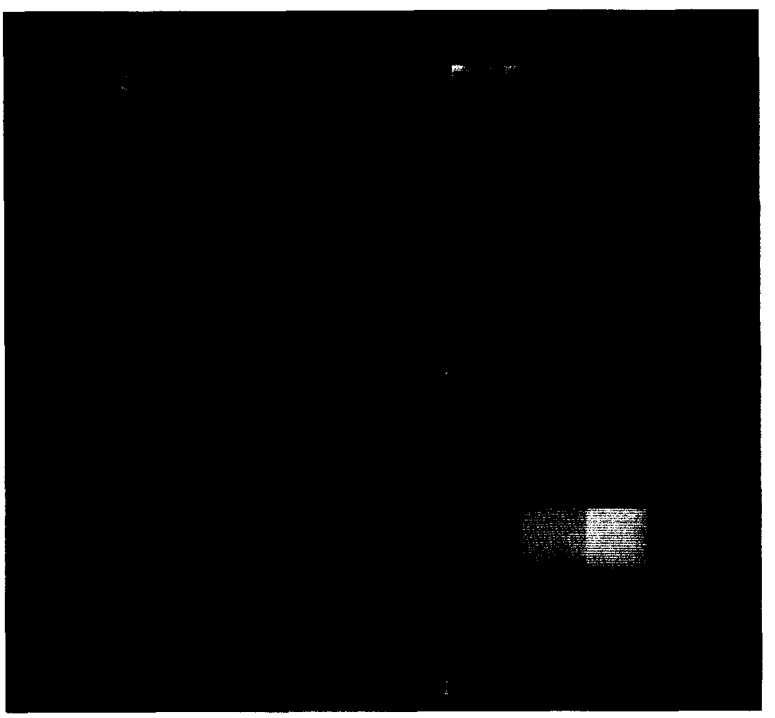

Fig. 10. Example 1, the selected area in the latent variable space is shown, in red, together with its representation in the image.

A more powerful tool is to use the scatter plot of principal components versus each other, also known as score plots. In a two-dimensional score plot, every single pixel is plotted, with two score values as its coordinates. A plot of the second component versus the first, for this example, shows that these components separate eight clusters of similar pixels (see Fig. 9). In this score plot we can simultaneously analyze all the revealed clusters and see what they represent in the original image. This is called latent variable space segmentation or feature space segmentation of an image. In Fig. 9 one of the classes was selected with a so-called score mask and Fig. 10 shows both the score mask and the classified area as red parts in the image. It is obvious that exclusively the quadratic sub-part from sample No. 7 is classified. The result is very clear. Proceeding with a full-feature space segmentation of all eight classes gave the result shown in Fig. 11. The percentage ratio of Co^{2+}/Ni^{2+} content for the classes that the clusters represent is given in the picture.

Although two of the samples were not separated by the first two components it is very clear that the mixture gradient of the two compounds

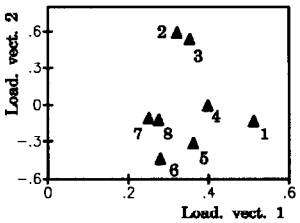

Fig. 12. Loading plot of the second loading vector versus the first for Example 1.

is caught by the MIA system. Especially the second component shows high correlation with the mixture gradient. Using the third component, all nine samples were classified apart.

The loading plot, Fig. 12, shows that all variables are positively correlated with the first component. Variables with a large influence of the informative second component are Nos. 2, 3 and 6 representing the wavelengths 498, 538 and 683 nm.

Besides being able to discriminate in the score plots, it is also possible to get other information from the score plot (Fig. 11). The size and shapes of the different clusters give an idea of random

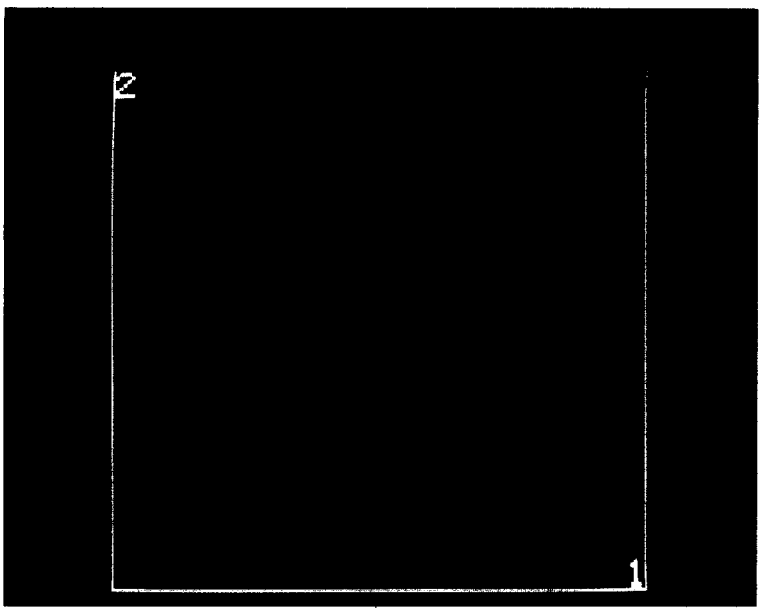

Fig. 11. Example 1, the eight separate score masks corresponding to the clusters. The percentage ratio of Co/Ni content for the classes that the clusters represent are given in the figure.

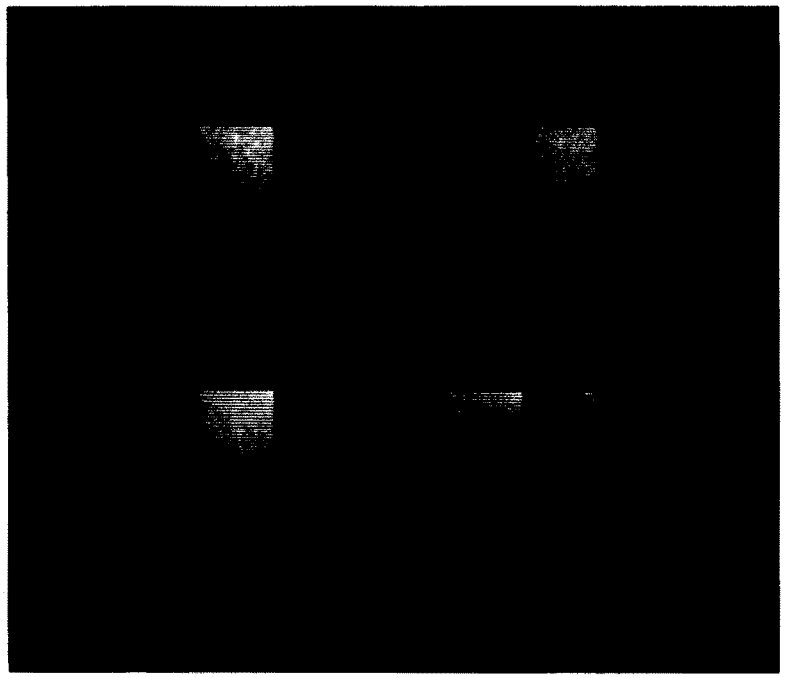

Fig. 13. Example 2, the first four images in the combined (150 × 117 × 8) multivariate raw image. Samples 1–6 are equal to the first example. Sample seven is a rectangular (150 × 17) sub-area on the bottom of the image. The wavelengths at which the images were generated, 459, 498, 538 and 583 nm, are given in the figure.

Fig. 14. Example 2, raw image Nos. 5–8 generated at 632, 683, 804 and 897 nm.

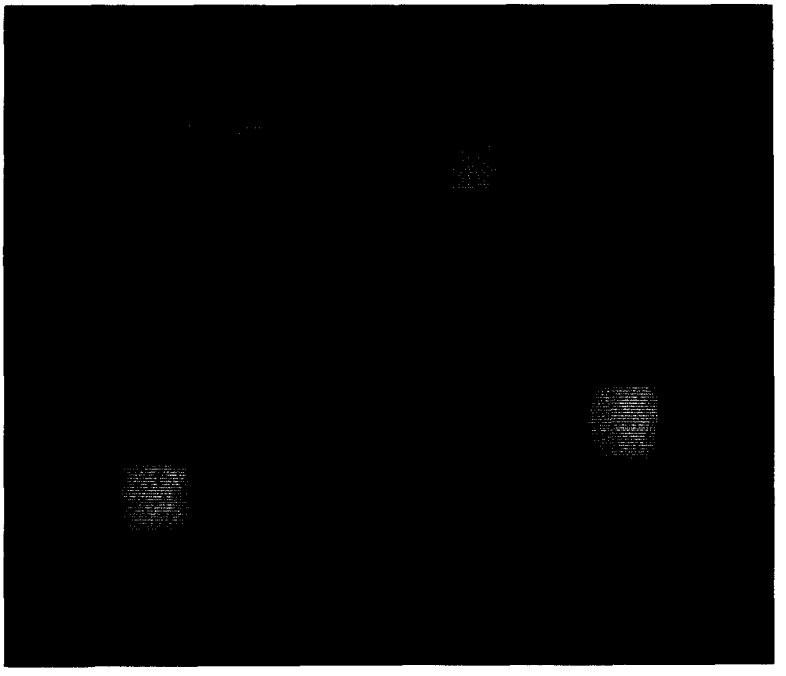

Fig. 15. Example 2, the four calculated principal component images.

and systematic error. The random error or natural variation inside a class is given by the size of the cluster for that class. The systematic errors for a class are given by oddities in the shape. The skew shape of the clusters, noticed in Fig. 12, is probably due to illumination error or an imper-

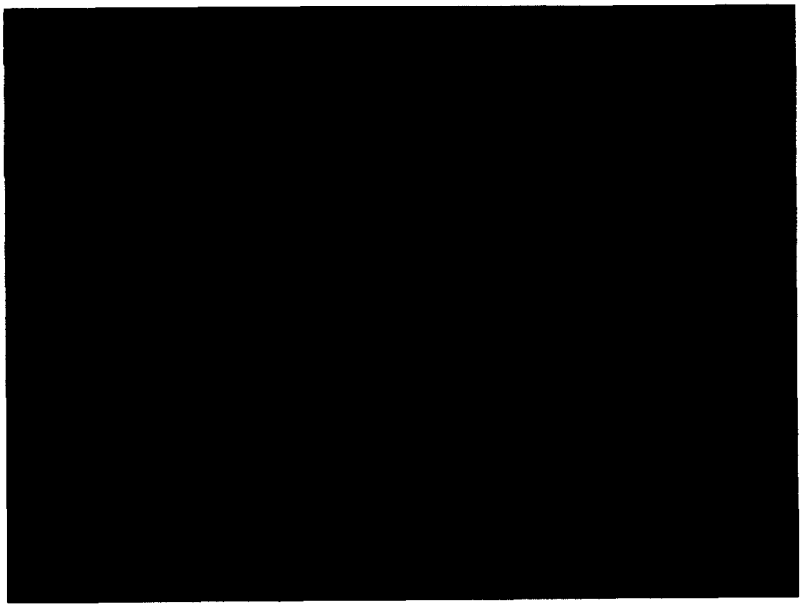

Fig. 16. Example 2, a RGB composite image. The second principal component in the blue channel, the third in the green channel and the fourth in the red channel.

fection in the camera tube. This feedback information of the instrumentation is very useful for further developments and improvements of the system.

Second example

In the second example, experimental complexity was increased by adding one compound, copper (Cu^{2+}), to the experiment. The scheme for the mixtures is shown in Table 1. The design for the mixtures of the three compounds was chosen deliberately to span the whole possible space of variations. This makes the interpretation of the results and identification of irregularities easier. The procedure and treatment of the samples and raw images are equal to the first example.

The eight images, Figs. 13 and 14, were pretreated by linear stretching and with mean centering of the variables. A PCA was calculated and the rank of the matrix was, by subjective judgement, estimated as four. The four-component model explained 99.2% of the total sum of squares. Explained sum of squares for each component is given in Table 2. The first component was not used basically because it mostly explains the illumination and camera errors. Fig. 15 gives the PC images and Fig. 16 shows a RGB composite image of components 2, 3 and 4.

TABLE 3

Histogram peak grey values for the clusters of seven samples in four principal components (Example 2)

Samples				Score histogram peak values			
No.	Co (%)	Ni (%)	Cu (%)	*t*1	*t*2	*t*3	*t*4
1	100	0	0	112	186	116	140
2	0	100	0	145	12	132	112
3	0	0	100	148	44	48	196
4	50	50	0	15	50	196	158
5	50	0	50	36	134	76	132
6	0	50	50	36	44	48	108
7	33	33	33	15	82	108	126

In Fig. 17, a traditional score plot is shown together with two intensity histogram profiles corresponding to the two score vectors. The histogram shows which intensities (grey values) dominate in the score images. The peak intensity values determined from the histogram are ap-

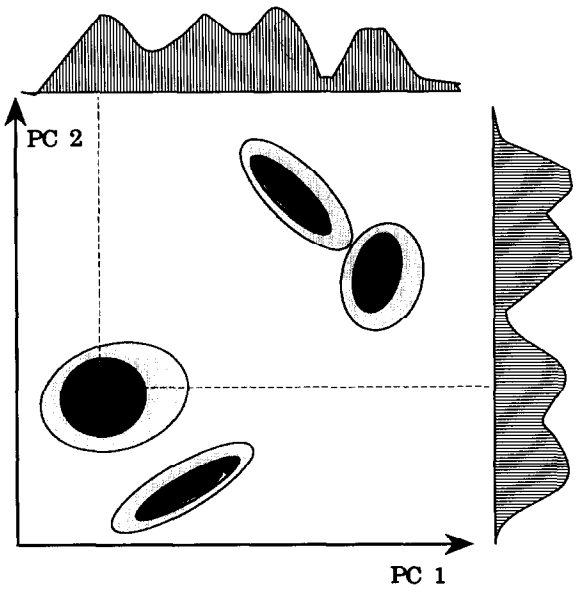

Fig. 17. The scatter plots of two score images (PC2 versus PC1). The objects are pixels and there are so many of them that they have to be represented by density functions. The density functions show clusters, gradients and outliers. Using score plots in combination with the grey value histograms, one can find the peak grey values of the clusters.

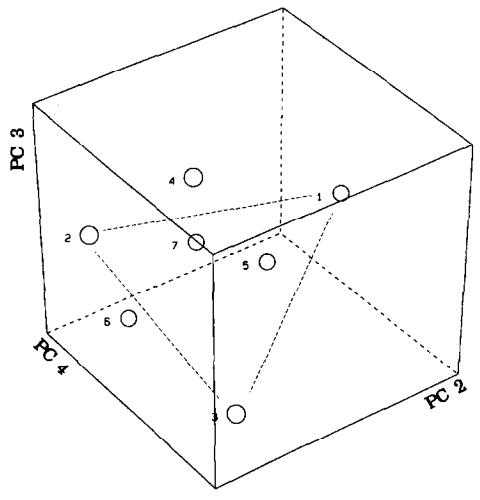

Fig. 18. A three-dimensional score plot for components 2, 3 and 4 (Example 2). The classes are represented by their histogram peak grey value for graphical representation. It is easy to see the design of the calibration samples represented in this plot, with a certain amount of distortion, caused by the measurement system errors and nonlinearities.

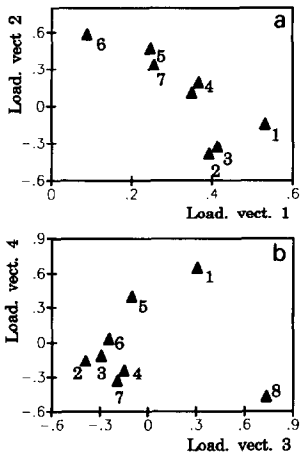

Fig. 19. (a) Loading plot of the second loading vector versus the first for the second example. (b) Loading plot of the fourth loading vector versus the third for the second example.

proximations of the score median values for appearing clusters in the score plot.

Using this technique for components 2, 3 and 4, Example 2, in combination with feature space segmentation, gives the results in Table 3. Plotting these seven peak grey values for the used component in a three-dimensional plot (Fig. 18), gives a typical pattern of a three-component mixture design. The pure compounds give the triangle corners, the 50/50 mixtures in between and the mixture of all three compounds is found in the middle. The small deformations of the design are due to nonlinearities and other errors in the measurement.

Fig. 19 shows the scatter plots of the loading vectors for all four components. Noticeable in Fig. 19b, is the strong influence of variable No. 1 (459 nm) on the fourth component and variable No. 8 (897 nm) on the third component. The interpretation of the loading plot, Fig. 19a, is fairly the same as for the first example.

SUMMARY AND CONCLUSIONS

Binary mixtures of Ni and Co (Example 1) and ternary mixtures of Ni, Co and Cu (Example 2) were measured as 512×512 images at eight wavelengths in the region 460–900 nm. The mixture compositions were constructed using a design, for ease of interpretation of the results. Uninteresting background and obvious errors by reflections of the source were eliminated by taking 50×50 samples from the 512×512 images. The $150 \times 150 \times 8$ multivariate image for Example 1 and the $150 \times 117 \times 8$ multivariate image for Example 2 were subjected to PCA after linear stretching and mean centering of the variables.

The results of the PC analyses are loading vectors and score images. These can be studied by making scatter plots: the score and loading plots. The loading plots were interpreted as importance of the different variables (wavelengths) in the models. The score plots showed clusters corresponding to the different mixture classes. In this study, binary and ternary mixtures of metal ions were completely classified using feature space segmentation in the latent variable space.

Encouraged by the satisfactory result further investigations will be made. In the future we will expand the complexity of the samples even more, dealing with mixtures of more than three compounds. The method can hopefully also be applied to a wider variety of substances.

It has been shown that important information concerning the instrumental setup is put forward by a PCA. In this study, the total concentration of the included substances is constant (100%) and mean centering of the data is done before PC calculations. For such a system, the PCA on the first data set (Example 1), should result in only one significant component describing the concentration difference between the samples. Example 2 should in an analogous manner be described by the first two principal components on that data set. As we have seen, this is not the case for the two examples. This means that additional information, besides the concentration differences of the samples, is present in the images. In this case, it seems likely that illumination errors in combination with a biased camera tube, have added additional structural information to the images. This feedback information of the equipment and the instrumental setup is very useful for further developments and improvements of the system.

The rank of the matrices was determined by step-wise viewing of the PC images for increasing dimensions of the model. When reaching higher

dimensions, the structure in the image disappears and just noise is displayed. This component together with the following components are rejected. Other, statistically more valid methods [26] exist but have not been implemented and applied to image data structures. Methods like, for example, cross-validation [26], are very computer-intensive. Using this method on very large data matrices, such as multivariate images (2 000 000 data points), will definitely create a slow, non-interactive system. New fast algorithms for effective cross-validation on large matrices must be developed before using these methods in multivariate image analysis.

The aim of the study was strictly an exploratory study of MIA, so no thorough investigation of the predictive capability of the models has been done. Dividing the data set into sub-parts, training set for model development and test set for model validation, is the most frequently used approach for model validation. Comparing predictions to true observed values for the test set gives a good measurement of the predictive capability of the derived model.

Algorithms for latent variable regression and calibration have recently been applied to multivariate image analysis. PCR has already been used with success on satellite images [27]. Work is going on for speeding up and simplifying PLS and PCR algorithms for large matrices.

Many industrial applications for univariate image analysis are already found. Especially in an industrial process there is a great need for quick automatic quality control. Image analysis systems can be developed for interactively analyzing products differing in intensity, size or shape. The products with wrong features are found and rejected. Introducing MIA into such a system provides the opportunity of also finding products of wrong chemical content.

ACKNOWLEDGEMENTS

Helén Bergner of the Center for Peat Research, Umeå, Sweden, is thanked for assistance in sample preparation and measurement. The Image Analysis section of the Swedish National Board for Technical Development is thanked for a grant. Lennart Eriksson, of the University of Umeå, is thanked for proofreading.

REFERENCES

1 W. Pratt, *Digital Image Processing*, Wiley, New York, 1978.
2 R. Gonzalez and P. Wintz, *Digital Image Processing*, Addison-Wesley, Reading, 2nd ed., 1987.
3 J. Lim, *Two-dimensional signal and image processing*, Prentice-Hall, Englewood Cliffs, 1990.
4 A. Rosenfeld and A. Kak, *Digital Picture Processing*, Academic Press, New York, 1982.
5 M. Horst, *Digital Image Processing of Remotely Sensed Data*, Academic Press, New York, 1987.
6 J. Jensen, *Introductory Digital Image Processing: A Remote Sensing Perspective*, Prentice-Hall, Englewood Cliffs, 1986.
7 T. Lillesand and R. Kiefer, *Remote Sensing and Image Interpretation*, Wiley, New York, 1987.
8 F. Sabins, *Remote Sensing Principles and Interpretation*, Freeman, San Francisco, 1978.
9 R. Barer and V. Coslett (Editors), *Advances in Optical and Electron Microscopy, Vol. 10*, Academic Press, London, 1987.
10 T. Mulvey and C. Sheppard (Editors), *Advances in Optical and Electron Microscopy, Vol. 11*, Academic Press, London, 1988.
11 J. Miller, Techniques for NMR imaging of solids, *Trac*, 10 (1991) 59–64.
12 F. Lindgren and P. Geladi, Multivariate image analysis: Exploring of spectral information of chemicals, *Proceedings from "Symposium i Anvendt Statistik", Copenhagen, 1991*, UNI.C, pp. 319–333.
13 P. Geladi, E. Bengtsson, K. Esbensen and H. Grahn, Image Analysis in Chemistry. Part 1: Properties of images, greylevel operations, the multivariate image, *Trac*, (1991) in press.
14 P. Geladi, K. Esbensen and S. Wold, Image analysis and chemical information in images, *Analytica Chimica Acta*, 191 (1986) 473–480.
15 P. Geladi and K. Esbensen, Can image analysis provide information useful in chemistry, *Journal of Chemometrics*, 3 (1989) 419–429.
16 K. Esbensen and P. Geladi, Strategy of multivariate image analysis (MIA), *Chemometrics and Intelligent Laboratory Systems*, 7 (1989) 67–86.
17 I.T. Jolliffe, *Principal Component Analysis*, Springer Verlag, New York, 1986.
18 H. Martens and T. Næs, *Multivariate Calibration*, Wiley, Chichester, 1989.
19 W.J. Dunn III, S. Wold, U. Edlund, S. Hellberg and J. Gasteiger, Multivariate structure–activity relationships between data from a battery of biological tests and an ensemble of structure descriptors: The PLS method,

Quantitative Structure–Activity Relationships, 3 (1984) 131–137.
20 S. Hellberg, *A Multivariate Approach to QSAR*, Thesis, University of Umeå, Umeå, 1986.
21 B. Skagerberg, *Principal Properties in Design and Structure Description in QSAR*, Thesis, University of Umeå, Umeå, 1989.
22 L. Erikssons, *A Strategy for Ranking Environmentally Occurring Chemicals*, Thesis, University of Umeå, Umeå, 1991.
23 S. Wold, K. Esbensen and P. Geladi, Principal component analysis, *Chemometrics and Intelligent Laboratory Systems*, 2 (1987) 37–52.
24 P. Geladi, H. Isaksson, L. Lindqvist, S. Wold and K. Esbensen, Principal component analysis on multivariate images, *Chemometrics and Intelligent Laboratory Systems*, 5 (1989) 209–220.
25 P. Geladi, Analysis of multi-way (multi-mode) data, *Chemometrics and Intelligent Laboratory Systems*, 7 (1989) 11–30.
26 S. Wold, Crossvalidatory estimation of the numbers of components in factor and principal component models, *Technometrics*, 20 (1978) 379–406.
27 P. Geladi and K. Esbensen, Regression on multivariate images: principal component regression for modeling, prediction and visual diagnostic tools, *Journal of Chemometrics*, 5 (1991) 97–111.